

The Cancer Imaging Archive (TCIA)

Janet F. Eary, M.D.

Associate Director, Cancer Imaging Program, National Cancer Institute

FNLAC, February 24, 2022



THE CANCER IMAGING ARCHIVE

Welcome to The Cancer Imaging Archive

The Cancer Imaging Archive (TCIA) is a service which de-identifies and hosts a large archive of medical images of cancer accessible for public download.

[SUBMIT YOUR DATA](#)

[ACCESS THE DATA](#)

TCIA : A Unique Public Resource for Cancer Research

THE **CANCER**
IMAGING ARCHIVE



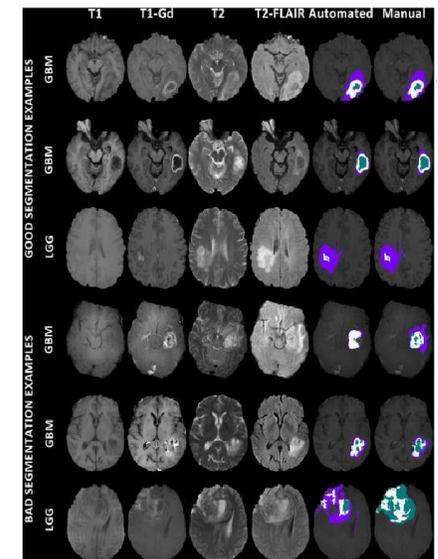
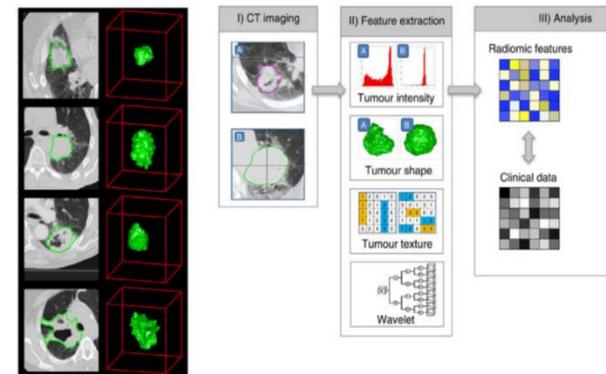
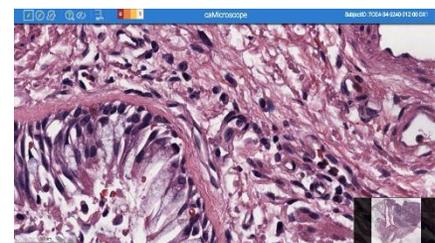
TCIA: NCI Open-access Image Data Sharing

- De-identified, curated image datasets
- Datasets for standard cancers, unique or rare types, specific clinical applications
- Emphasis on associated metadata for integrated data analyses
- Explores new image data types and develops procedures for sharing

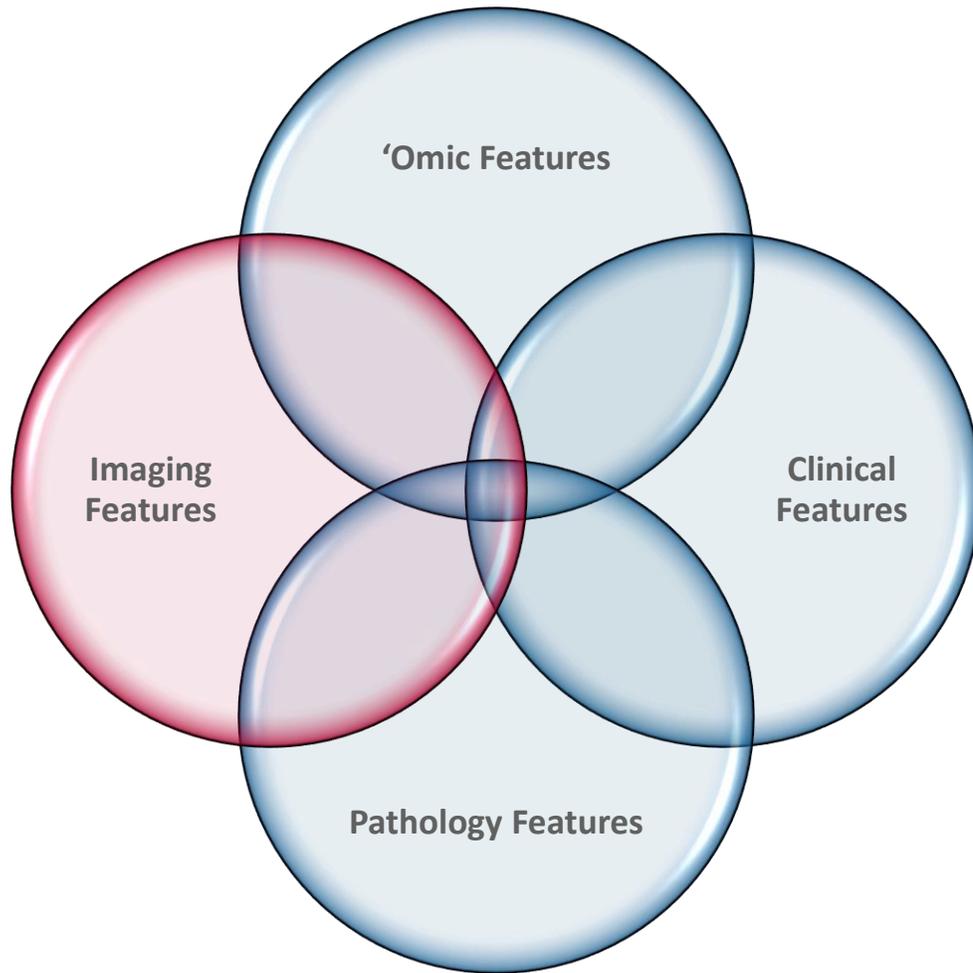
SCIENTIFIC DATA a nature research journal

Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani & Christos Davatzikos



TCIA addresses a critical research need



- **Images are very rich datasets**
- Use of imaging can inform cancer discoveries: providing
 - Temporal Information
 - Spatial Information
 - Useful in integration with other biomarkers
 - Non invasive

Funding, Governance, Implementation

➤ Funding and Governance



➤ Management and implementation

Frederick National Laboratory
for Cancer Research

John Freyman team lead



➤ Subcontract implementation

TCIA image collection sources

- Incoming proposals evaluated monthly by TCIA Advisory Group
- Expert reviewers from NCI Cancer Imaging Program, Cancer Diagnosis Program, Center for Cancer Research and Frederick National Lab

proposals
reviewed monthly



Data generated
by NCI/NIH Grants



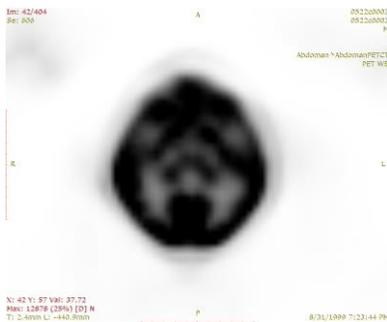
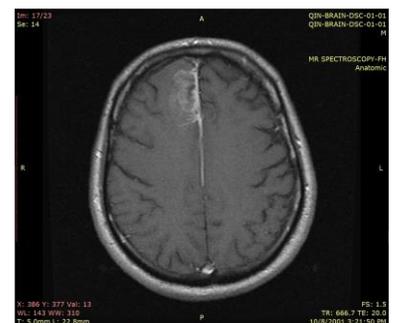
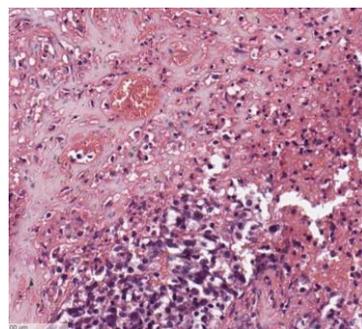
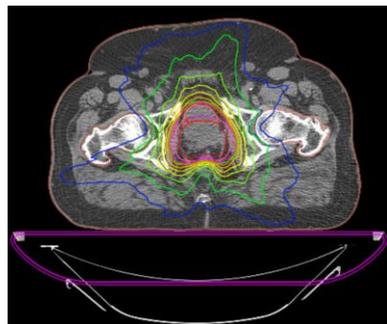
Challenge
competitions

SCIENTIFIC
DATA



Publication
data sharing
requests

TCIA Supports Radiology and Histopathology Imaging in Major NCI Research Initiatives



NIH **THE CANCER GENOME ATLAS**
National Cancer Institute
National Human Genome Research Institute

CLINICAL PROTEOMIC
TUMOR ANALYSIS CONSORTIUM



**Quantitative
Imaging
Network**

NCI **National Clinical
Trials Network**
a National Cancer Institute program

PDMR **NCI Patient-Derived Models Repository**
An NCI Precision Oncology InitiativeSM Resource

National Lung Screening Trial

ECOG-ACRIN
cancer research group
Reshaping the future of patient care

IROC[®]
IMAGING AND
RADIATION ONCOLOGY CORE
Global Leaders in Clinical Trial Quality Assurance

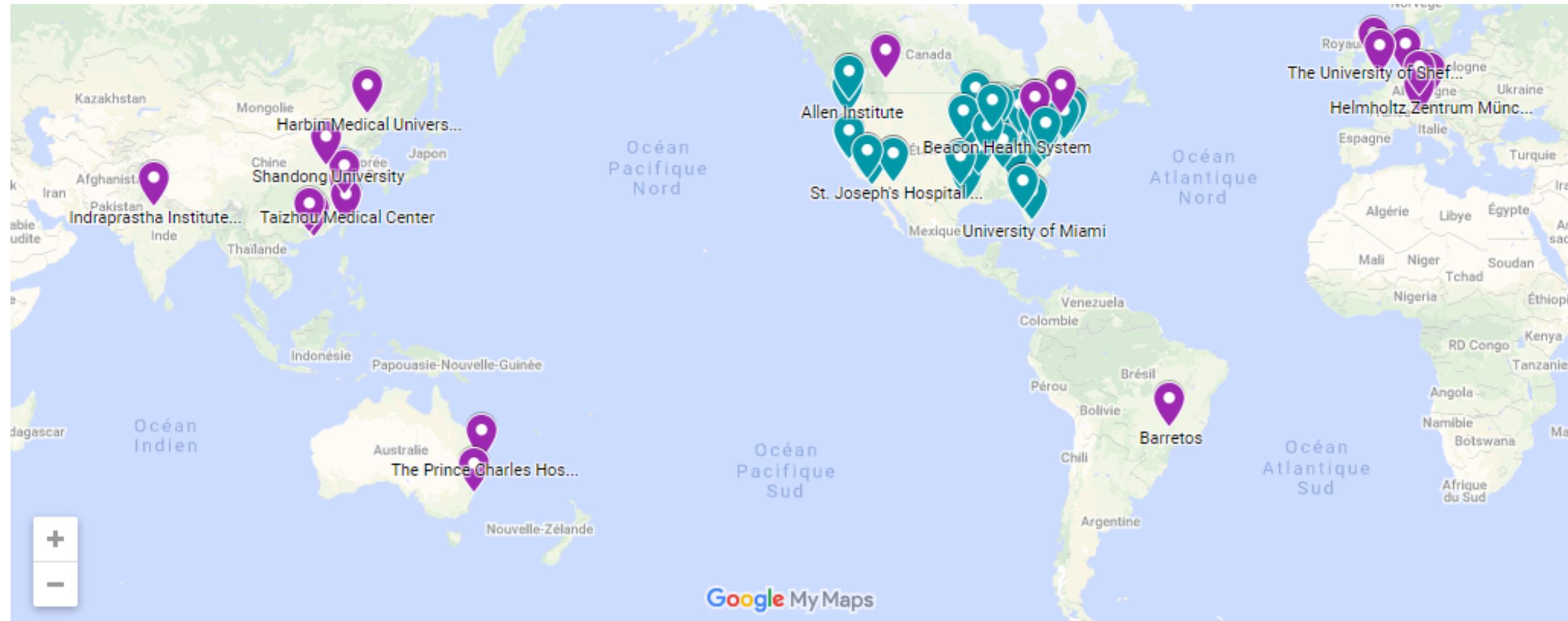
Community
Initiated

Image de-identification and curation into well described datasets

TCIA Advisory Group – Review Criteria

- How important is this data set to facilitating **research**?
- Does this dataset address a **data gap** for current research or a clinical need?
- Is this a **novel/unique** dataset compared to what's already in TCIA?
- Is the dataset of a sufficient **size/scale** to support research/analyses or hypothesis development?
- Does the dataset contain sufficient **supporting data and documentation**?
- If the dataset consists of an analysis of image based data, is it **based on a biological hypothesis*** or other proposed discovery about the patho-physiological basis of cancer?
 - How will other **researchers benefit** if this data is hosted by TCIA?
 - What scientific criteria was used to determine **the methodology of image analyses**?

85+ Contributing Institutions to TCIA



TCIA components



Data Collection

Specialized de-identification processes and tools support data collection, curation, and de-identification

Specialized Teams

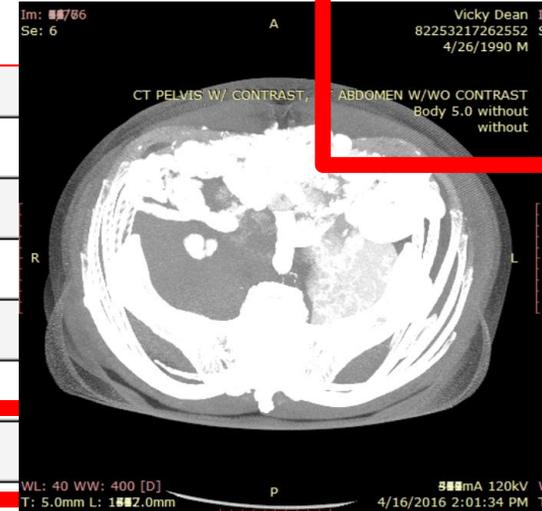
Data Hosting and Query

Research-focused website GUI and programmatic interfaces

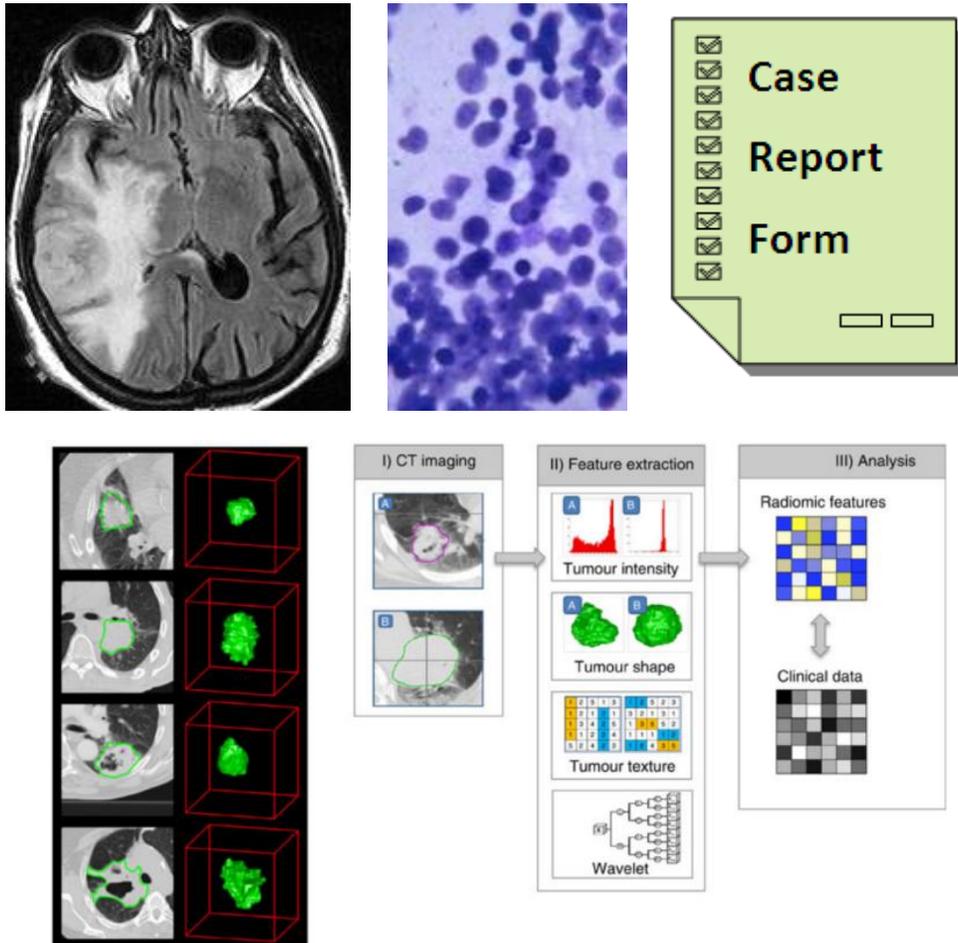
Sharing imaging data presents challenges

- PHI can appear in hundreds of places in images
 - Pixel data
 - Dates
 - Identifiers
 - Descriptions
 - “Private tags” with ambiguous proprietary content
 - Content that can be used for identification of an individual

(0008,0022)	Acquisition Date	19970614
(0008,0023)	Content Date	19970614
(0008,0024)	Study Time	073837
(0008,0031)	Series Time	074251
(0008,0032)	Acquisition Time	074252
(0008,0033)	Content Time	074252
(0008,0050)	Accession Number	2819497684894126
(0008,0060)	Modality	MR
(0008,0070)	Manufacturer	GE MEDICAL SYSTEMS
(0008,0090)	Referring Physician's Name	
(0008,1010)	Station Name	
(0008,1030)	Study Description	MRI, BRAIN W&W/O CONTRAMR
(0008,1032)	Procedure Code Sequence	
(0008,103E)	Series Description	AX FLAIR
(0008,1090)	Manufacturer's Model Name	GENESIS_SIGNA
(0010,0010)	Patient's Name	
(0010,0020)	Patient ID	TCGA-02-0009
(0010,0030)	Patient's Birth	



TCIA Data



- 160 Collections consisting of ~50,000 subjects with over 50 million images
 - 6 murine collections (4 PDMR models)
 - Canine GBM pilot dataset from ICDC
- 29 Image Analysis Result datasets based on TCIA collections
- Radiology, radiation therapy, and histopathology images
- variety of cancer histologies
- Most collections have associated supporting data
 - DICOM data elements
 - Demographics/outcomes/therapy
 - Image analyses and results
 - Genomics/Proteomics
- COVID-19 initiative

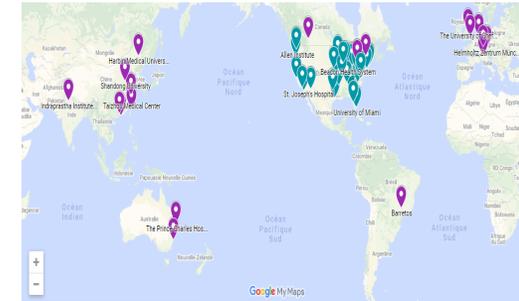
TCIA Scientific Impact

Data accessed from 216 countries

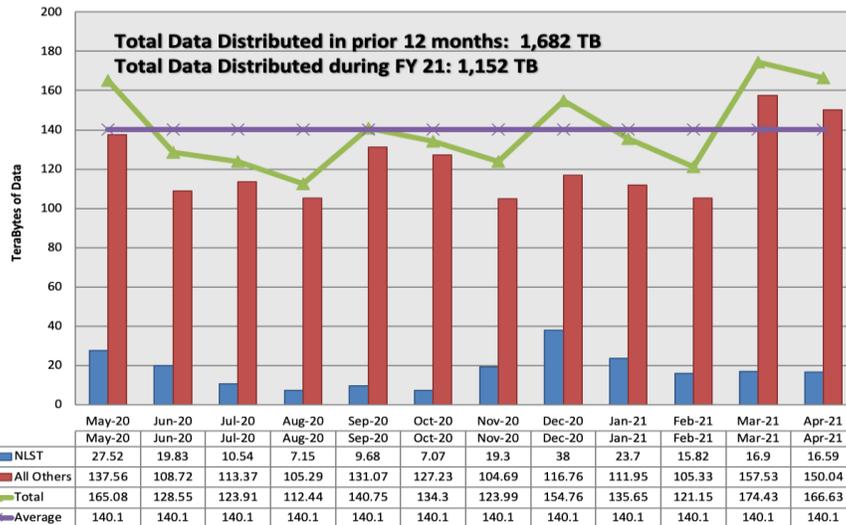


- 768,000 users from 216 countries and regions
- 20K users access TCIA each month
- average downloads of 140TB of data per month, nearly 1.7PB per year
- data publisher and repository of record for **Nature**, **PLOS One**, **Medical Physics**, **Elsevier**, **F1000 Research**,...

85+ Contributing Institutions

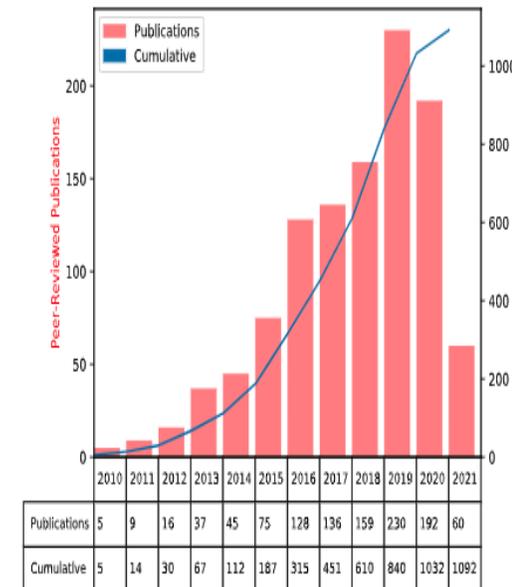


TCIA Data Distribution to Researchers



- Data collected from over 85 institutions
- Over 1,300 peer-reviewed publications

Publications Over Time



TCIA is leading the way for public data sharing

TCIA:

- Collects ORCID for unambiguous author attribution
- Assigns Digital Object Identifiers (DOIs) to each dataset
 - Ensures unambiguous dataset attribution
 - Enables authors to receive academic credit through data citations
- Publishes data using Creative Commons licensing to ensure clear data usage policies that permit research and commercialization
- Avoids unnecessary account creation or login requirements for open-access datasets
- Partners with journal publishers to provide a data sharing repository for data sharing requirements

NIH has recognized TCIA as an NIH High Value Data Asset (HVDA)

nature

COMMENT · 04 JUNE 2019 · CORRECTION 05 JUNE 2019

Credit data generators for data reuse

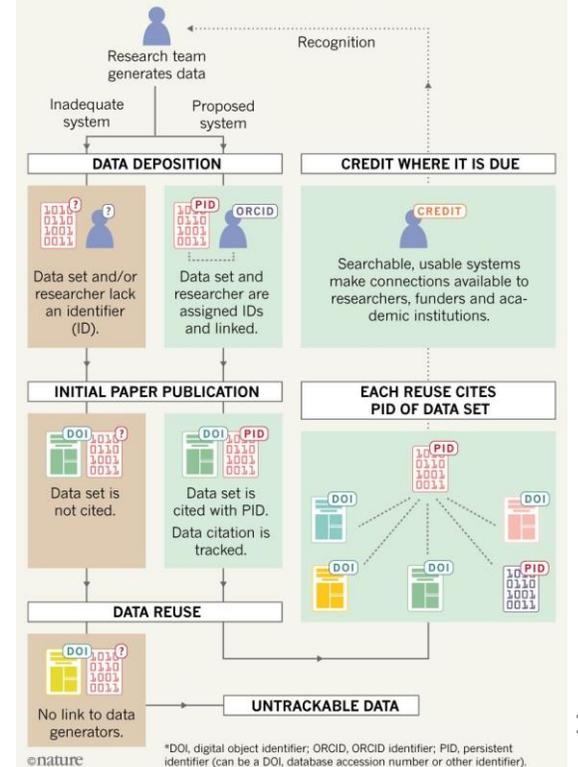
To promote effective sharing, we must create an enduring link between the people who generate data and its future uses, urge Heather H. Pierce and colleagues.

Heather H. Pierce, Anurupa Dev, Emily Statham & Barbara E. Bierer

VIRTUOUS CYCLE

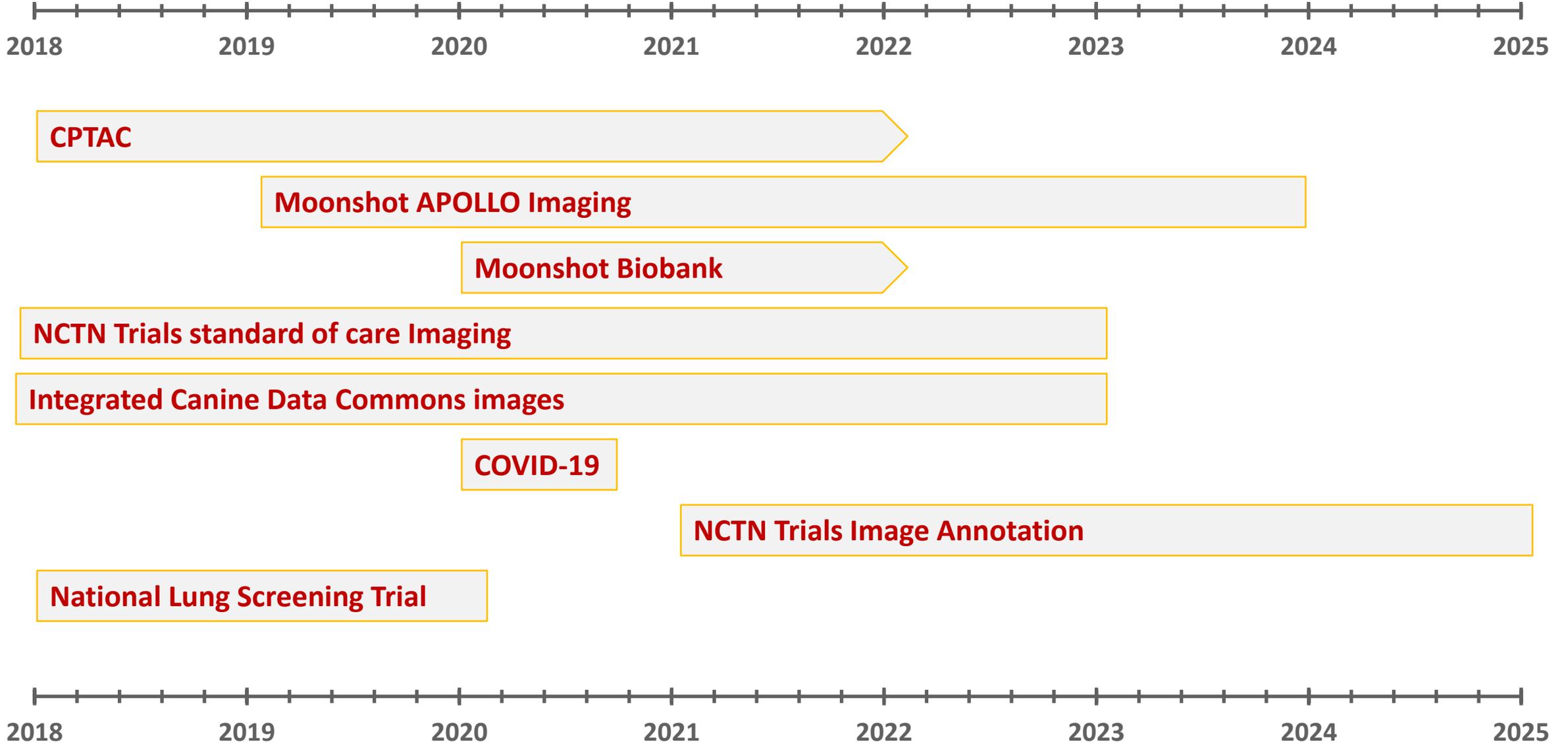
Linking people to the data they generate will lead to ways to credit them when data are reused. This would influence funding and promotion, and incentivize more (and better) curation and sharing.

PID Data set ID*
ORCID Researcher ID
DOI Paper ID



Supporting NCI Programs:

NCI has leveraged TCIA to add imaging data value to existing programs



Clinical Proteomic Tumor Analysis Consortium: TCIA Portal: Connecting all data types

TCIA collects, de-identifies and hosts standard-of-care imaging from CPTAC patients

TCIA links radiology and histopathology imaging to the CPTAC analysis results on the Genomics and Proteomics Data Commons.

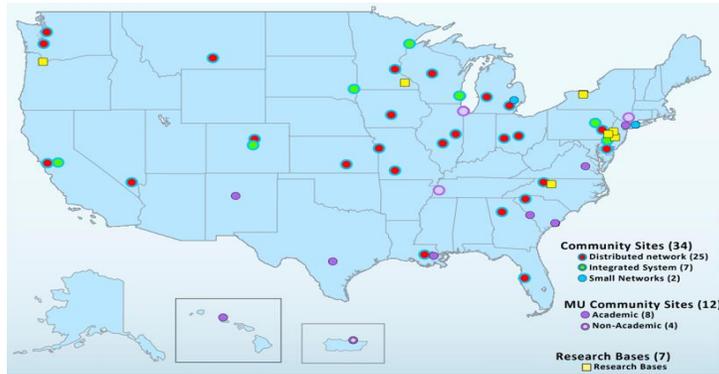
Case ID	Project	Primary Site	Gender	Files	Available Files per Data Category					
				Seq	Exp	BNV	CNV	Meth	Clinical	Blk
C3L-00016	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00016-21	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00016-24	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00019	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00019-23	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00019-24	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00025	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00025-21	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00025-22	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00025-23	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00278	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00278-23	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00278-21	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00278-22	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00349	CPTAC-3	Brain	Male	25	2	10	0	0	0	0
C3L-00349-23	CPTAC-3	Brain	Male	25	2	10	0	0	0	0

Moonshot APOLLO Clinical Network



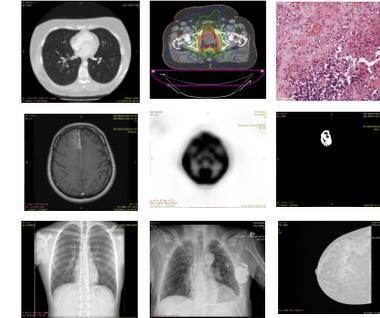
- Trilateral Collaboration –   
- Genomic and proteomic screening to target therapies for up to 8,000 active service and veterans.
- TCIA is establishing first-of-its kind clinical imaging de-identification and sharing systems within VA and DOD.
- Imaging from DOD, VA and Civilian sites will be posted on TCIA and mined for clinically-relevant information in combination with APOLLO proteogenomic findings.

Cancer MoonshotSM Biobank

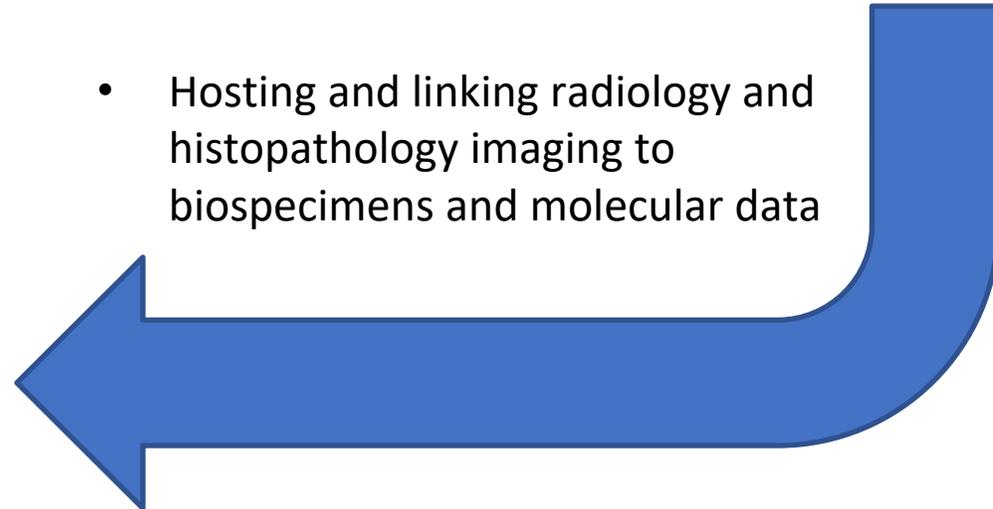


- 2000 patients from NCI Community Oncology Research Program (NCORP)
- Longitudinal data from full course of Treatment
- Biospecimen Processing and Storage, Data Quality Assurance
- Molecular Characterization and PDX models
- CLIA Tumor Characterization Assays and Aggregate research results provided to the community

Collaboration with IROC for Collection of Radiology imaging at multiple timepoints for each patient

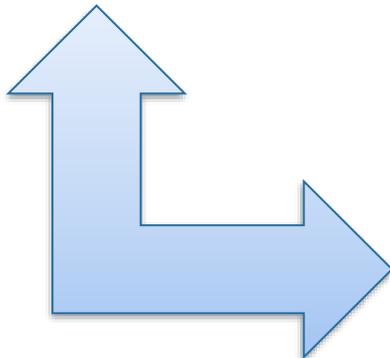


- Hosting and linking radiology and histopathology imaging to biospecimens and molecular data



Imaging in NCI Clinical Trials

Imaging from selected NCTN Trials are being collected, curated and posted on TCIA with links to the clinical data on the NCTN Data Archive



NATIONAL CANCER INSTITUTE
NCTN/NCORP Data Archive

AREN0533	NCT00379340
AREN0534	NCT00945009
ACNS0332	NCT00392327
AHEP0731	NCT00980460
CALGB50303	NCT00118209
ACRIN-Contralateral-Breast-MR (ACRIN 6667)	NCT00058058
ACRIN-HNSCC-FDG-PET/CT (ACRIN 6685)	NCT00983697
ACRIN-DSC-MR-Brain (ACRIN 6677/RTOG 0625)	NCT00433381
NSCLC-Cetuximab (RTOG-0617)	NCT00533949
ACRIN-NSCLC-FDG-PET (ACRIN 6668)	NCT00083083
ACRIN-FMISO-Brain (ACRIN 6684)	NCT00902577
ACRIN-FLT-Breast (ACRIN 6688)	NCT00572728
ISPY1 (ACRIN 6657)	NCT00043017
NRG-1308 (RTOG 1308)	NCT01993810
National Lung Screening Trial	NCT00047385
Head-Neck Cetuximab (RTOG 0522)	NCT00265941
CT Colonography (ACRIN 6664)	NCT00084929

Integrated Canine Data Commons

- TCIA is partnering with ICDC to curate and host radiology and pathology imaging data



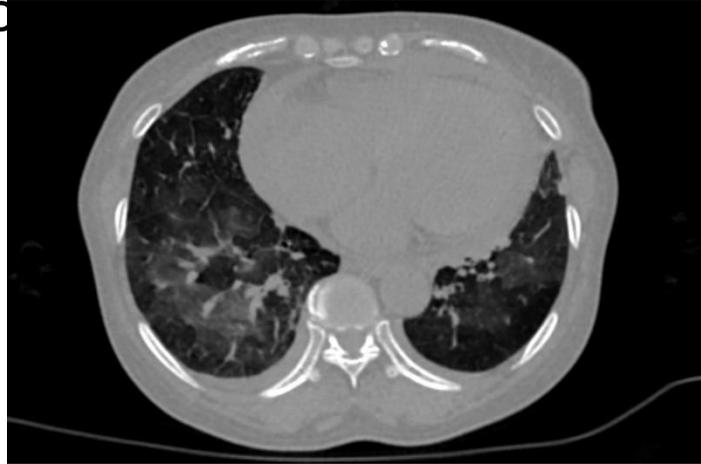
ICDC-Glioma01

This is the first dataset of its kind to comprehensively describe and report the clinical, pathologic, imaging and genomic landscape of naturally-occurring canine glioma.



TCIA COVID-19 Response

- Imaging has been an important way to characterize COVID-19. Making public the appearance and phases of the disease
- TCIA quickly moved to hosting and sharing COVID-19 data sets with its established process to handle large volumes of radiological imaging data
- TCIA played a critical role in the launching of NIBIB's Medical Imaging Data Resource Center (MIDRC) where COVID-19 imaging is the first use case

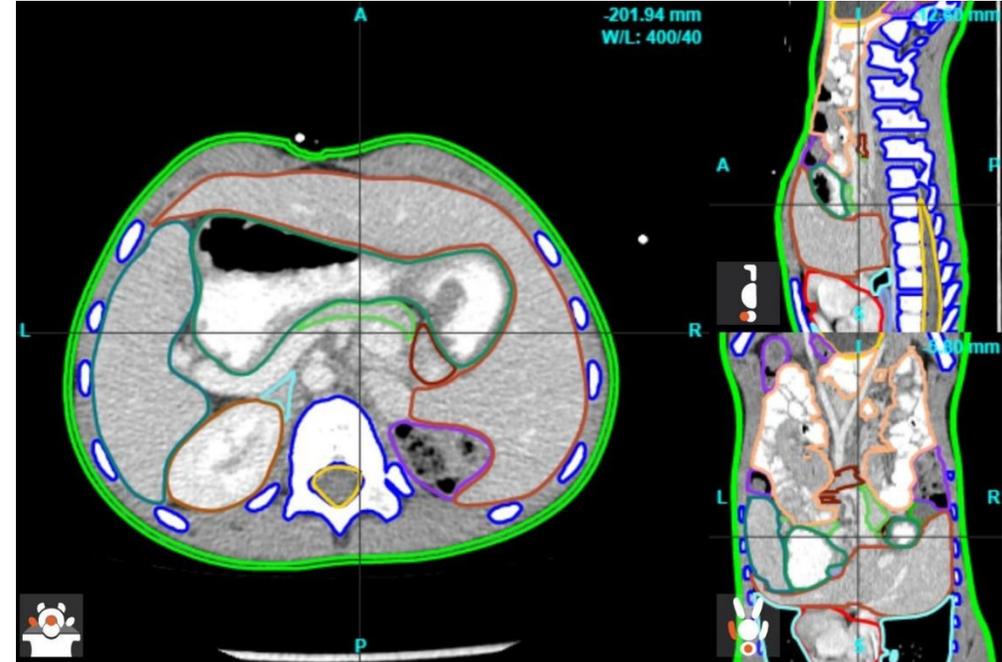


Datasets Published in TCIA

1. Chest Imaging with Clinical and Genomic Correlates Representing a Rural COVID-19 Positive Population
2. CT Images in COVID-19
3. MIDRC – RICORD Chest CT Covid Positive
4. MIDRC – RICORD Chest CT Covid Negative
5. MIDRC – RICORD Chest Chest x-ray Covid Positive
6. Stony Brook University COVID-19 Positive Cases

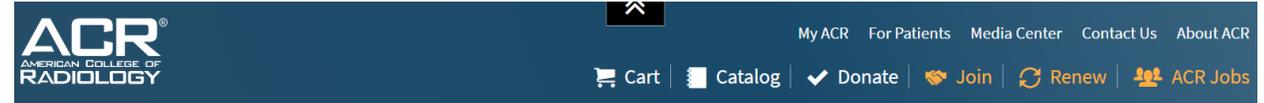
Childhood Cancer Data Initiative

- TCIA supports CCDI goals
 - 4 Childrens Oncology Group clinical trial datasets currently available with more under curation
 - 1 community dataset published
 - Engaging with CCDI Data Catalog team to ensure TCIA datasets are discoverable for investigators



TCIA prepares data for Artificial Intelligence Research

- NCTN Clinical Trial Annotation project
- Linking ACR Clinical Use Cases to TCIA Collections
- Connecting to major medical imaging AI platforms
- Providing highly curated data to AI-focused challenge competitions



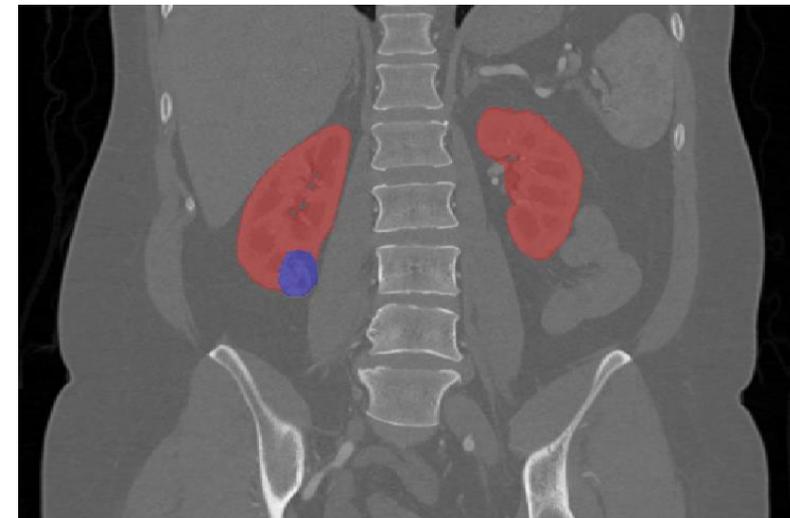
April 26, 2021

ACR Data Science Institute Links Use Cases to NCI Archive to Speed AI Development

Innovative Approach Links Datasets to Use Cases to Provide a One-stop Resource for AI Developers

The American College of Radiology® (ACR®) [Data Science Institute](#)® (DSI) and the [Cancer Imaging Archive](#)® (TCIA), funded by the National Cancer Institute (NCI), have teamed up to connect use cases and datasets to speed medical imaging artificial intelligence (AI) development. Dataset collection is the most important step in developing robust AI algorithms. Linking ACR DSI Define-AI use cases to datasets enables developers to build radiology AI algorithms that include defined data elements and DICOM images useful for inputs, outputs, and training and testing models.

TCIA datasets have been matched to ACR DSI cancer and non-cancer use cases based upon attributes such as body area, modality and presence of secondary comorbidities. TCIA data are available under Creative Commons Attribution Licenses, and most are freely available for commercial use for machine learning purposes.

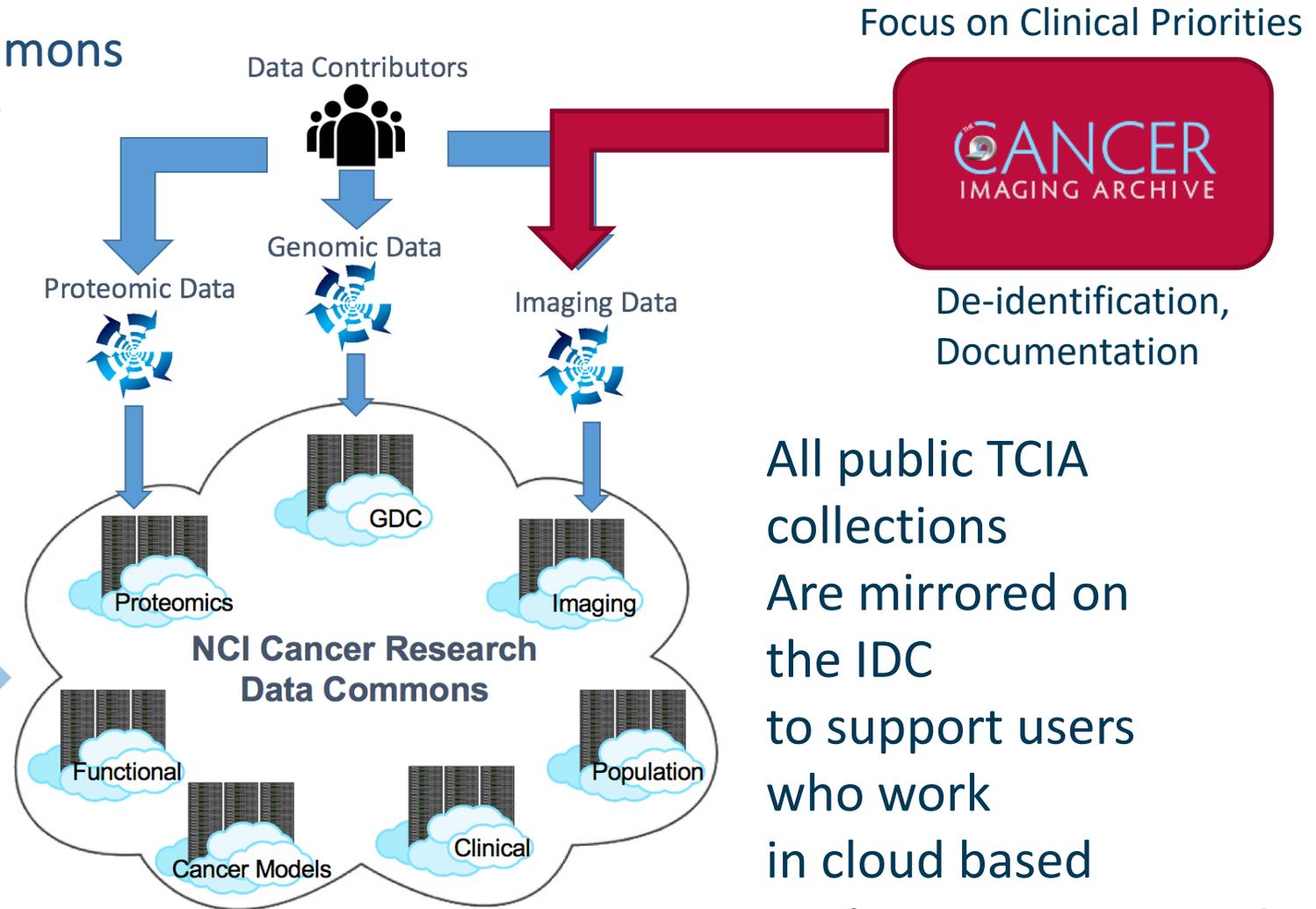
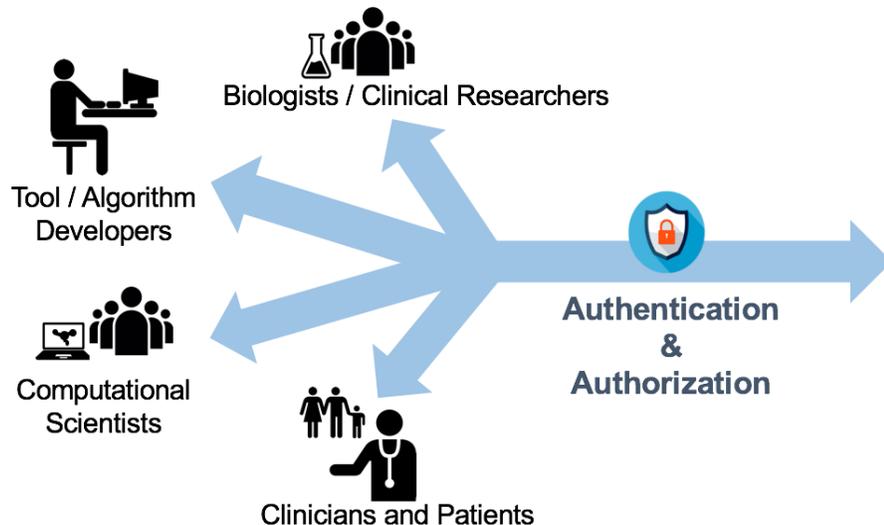


TCIA data in the NCI Imaging Data Commons for AI

The NCI Cancer Research Data Commons A virtual, expandable infrastructure

- Standardized data submission and Q/C
- Controlled vocabularies
- Harmonization by subject matter experts

- Secure data access through API or web UI
- Query across data domains
- Analytics, elastic compute, visualization



All public TCIA collections
Are mirrored on the IDC
to support users who work
in cloud based environments

Acknowledgements

NCI Cancer Imaging Program

Janet F. Eary, M.D.

Lalitha K. Shankar, M.D., Ph.D.

Fredrick National Laboratory for Cancer Research

John Freymann Team Lead

Justin Kirby

Brenda Fevrier-Sullivan

Pam Angelus

Luis Cordeiro

Michelle Tacconelli

NBIA Team (NCI/FNLCR/Ellumen)

Ulli Wagner

Scott Gustafson

Qinyan Pan

Russ Rieling

Carolyn Klinger

Martin Lerner

Tin Tran

Keyvan Farahani

University of Arkansas

Dr. Fred Prior

Kirk Smith

Lawrence Tarbox

Dwayne Dobbins

Jeff Tobler

Sonya Utecht

Julie Frund

Bill Bennett

Tracy Nolan

Quasar Jarosz

Brittney Camp

Erica Bilello

Claren Freeman

Natasha Honomichl

Geri Blake

Diana Stockton

Erica Bilello

Robert Brown

Emory University

Dr. Ashish Sharma

Ryan Birmingham

Contractors

John Perry

Betty Levine

Imaging and Radiation

Oncology Core

Dr. Michael Knopp

Connor Nealer

THANKS!